Copyright © 2002, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

USING DUAL-SUPPLY, DUAL-THRESHOLD AND TRANSISTOR SIZING TO REDUCE POWER IN DIGITAL INTEGRATED CIRCUITS

by

Stephanie Ann Augsburger

Memorandum No. UCB/ERL M02/6

5 April 2002

COVER

USING DUAL-SUPPLY, DUAL-THRESHOLD AND TRANSISTOR SIZING TO REDUCE POWER IN DIGITAL INTEGRATED CIRCUITS

by

Stephanie Ann Augsburger

Memorandum No. UCB/ERL M02/6

5 April 2002

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Using Dual-Supply, Dual-Threshold and Transistor Sizing to Reduce Power in Digital Integrated Circuits

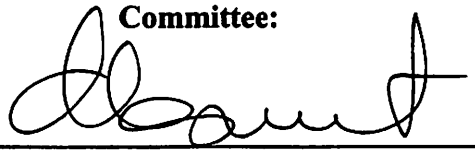
by Stephanie Ann Augsburger

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II.**

Approval for the Report and Comprehensive Examination:

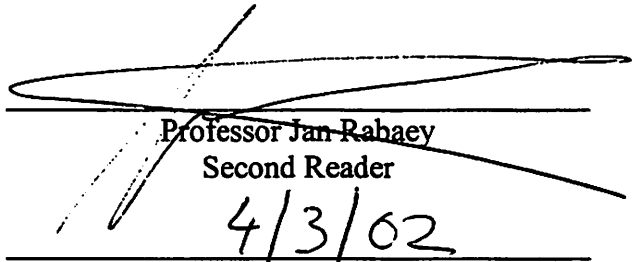
Committee:



Professor Borivoje Nikolić
Research Advisor

4/3/02
(Date)

* * * * * * *



Professor Jan Rabaey
Second Reader

4/3/02
(Date)

Abstract

In both high and low performance circuits, both active and leakage control of power consumption is critical. This research focuses on the use of multiple threshold voltages, multiple supply voltages and transistor sizing to reduce power in digital circuits. Non-critical paths are slowed down by either raising the threshold voltage, lowering the supply, downsizing gates, or a combination of these techniques.

A framework, based on gate models extrapolated from circuit-level simulation, was developed in order to evaluate these techniques. Using the framework, the effects of dual-supply (with two different values for the low supply voltage), dual-threshold and sizing were considered on a general logic block in order to gain a consistent idea of how and when these techniques should be used. In total, fifteen different techniques or combination of techniques were applied to the baseline design with varying results

This research shows promising results. Energy savings from these three base techniques can be compounded through proper combination for additional benefit. The order of application of these techniques determines the final savings in active and leakage power. Lowering supply, downsizing gates, and then raising transistor threshold, in order of effectiveness, are the keys to controlling active power. Multiple-threshold design is the most effective for leakage power control. It is believed that these results will motivate additional CAD support for designs employing a combination of power reduction methods.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Design techniques for low power consumption in modern VLSI are becoming increasingly important [Meindl95], [Chandrakasan92]. As technology moves into deep submicron feature sizes, power dissipation due to leakage current is increasing at an alarming rate. Projections show that leakage power will become comparable to dynamic power dissipation in the next few years [De00]. Dynamic power is also increasing, and still dominates. Supply voltage has not been scaled aggressively enough to keep power per unit area constant over technology generations. Exacerbating this problem is the growth in die area [Edmonson00].

One result of this increasing power in high-performance designs is the need for more sophisticated packaging with better thermal properties, leading to increased component cost. Large supply currents and *dI/dt* drops are another concern. Also, the need for low-power, high-performance VLSI is further fueled by the growing demand for portable devices such as cellular phones, laptops and PDA's. For such battery-operated devices, power consumption is paramount, and performance must somehow be maintained while decreasing power and hence increasing battery life. Leakage current (and power) has increased importance with burst mode type of computation. In this computational model, the majority of time is spent in idle mode. Large leakage currents during this sleep mode can cause battery drain in a portable system, and even in a non-portable system, is extremely wasteful.

Historically, VLSI's have been designed with a single supply voltage and a single threshold voltage. Process scaling was the primary mechanism by which the exponential growth in integration and performance was realized. While this scaling allowed enormous gains in operating frequencies, transistor count and performance, breakdown mechanisms, power consumption and reliability issues forced the supply voltage to be scaled with decreasing feature size. This in turn required threshold voltage scaling in order to maintain performance. This has a dramatic effect on leakage current, as subthreshold current increases exponentially with reduced threshold voltage.

Recently, however, multiple threshold voltages and multiple supply voltages have been used in an effort to reduce power consumption. Dual-threshold processes have become quite commonplace, starting with the 0.18μm generation, allowing designers to assign the low-threshold devices to critical paths and utilize high-threshold devices for non-critical paths. Alternatively, some experimental designs are now being reported which utilize dual (or triple) supply voltages.

Thus, the overall goal with low-power design is to identify any slack timing available, then eliminate this slack timing while saving power through the use of lowered supply voltage, increased threshold voltage, or smaller transistor sizes. Typically, this is achieved through the use of either a dual-V_T process or dual-V_{DD} design, along with sizing. However, there is nothing fundamentally limiting the designer's ability to use any number of threshold voltages or supply voltages in any arbitrary combination with or without transistor sizing. In practice, power savings must be weighed against increased manufacturing costs and design complexity when optimizing both the process and circuit design.

1.1 Scope of This Work

This work considers the effectiveness of the techniques of multiple threshold voltages, multiple supply voltages and sizing and the combination of these techniques on both active and leakage power reduction. The goal of this work is to determine the optimal design point (and if such an optimum exists). A secondary goal is to motivate CAD support for designs combining these techniques, particularly support for dual- or multi-

supply voltage designs. This work is restricted to constant throughput, constant latency logic blocks. For a given delay, power and energy are minimized. The target application for these results is general ASIC design, which is characterized by a limited number of paths that constitute the critical delay. The observation of cycle slack in the non-critical paths permits the introductions of these power reduction techniques without affecting the overall system throughput.

Early on in this project, it was determined that adequate CAD software support was not available to easily examine the effects of multiple threshold voltages, sizing, and especially multiple supply voltages. In order to facilitate this exploration, a design framework was constructed using MS Excel software and Spectre simulations. Models of basic gates were derived through simulation and then a generic path-delay distribution was generated based on these gates inside the design framework. Inside this environment, combinations of the three power reduction techniques were evaluated. Level-conversion penalties were considered with dual-V_{DD} designs. Two values were considered for the lower supply voltage in the dual-V_{DD} designs.

1.2 Related Work

In recent years, a number of power reduction techniques have emerged. These focused primarily on the individual effects of multiple-supply, multiple-threshold and transistor sizing techniques. In [Usami95], the basics of dual-supply design are explained. A 10% to 20% power reduction was reported with a clustered voltage scaling (CVS) dual-V_{DD} design. Dual-supply methodology, layout issues and such are covered in [Usami00]. [Kato00] presents the use of multiple-threshold assignment on a cell-by-cell basis and reports leakage reduction from 75% to 90%. A triple-threshold RISC processor was showcased in [Yamashita00]. The use of transistor sizing for power reduction is a common technique and has been covered thoroughly in years past [Rabaey96]. In [Hamada01], multiple-supply, multiple-threshold and transistor sizing were looked at from a theoretical standpoint. Rules of thumb were derived from a series of equations. However, little work has been done in terms of combing multiple-supply, multiple-threshold and sizing techniques in order to compound power savings and understanding if it works at all.

Other methods of power reduction are listed here but are not covered in this analysis. Supply voltage can be lowered either statically or dynamically. With a statically reduced supply voltage, throughput can be maintained by introducing parallelism or pipelining into the system [Chandrakasan92]. With dynamic voltage scaling (DVS), supply voltage is varied, and hence the frequency, in response to the computational load [Burd00]. In regards to leakage power, either active or standby leakage power can be attacked. Similar to DVS, threshold voltages can be adjusted based on computational load by changing the body bias. In [Miyazaki02], This technique was implemented in conjunction with DVS in a multiply-accumulate unit. MTCMOS, or multi-threshold CMOS, is a method to reduce standby leakage. High-V_T transistors are used to gate the power supplies to a low-V_T logic block. Sizing of the sleep transistors is critical to performance as discussed in [Kao00].

1.3 Thesis Overview

The remainder of this report describes the work undertaken to assess the effectiveness of multiple threshold voltages, multiple supply voltages and sizing. Chapter 2 provides a background on the principles of power in VLSI, including multiple supply and multiple threshold voltages, then Chapter 3 presents a theoretical study into the use of these techniques. Chapter 4 discusses the design framework designed for the evaluation of these techniques. This includes a preliminary study completed first, as well as the gate models used in the framework and an overview of level-converting flip-flops. Chapter 5 gives an in depth analysis into the results of this evaluation. Chapter 6 summarizes and concludes the report.

CHAPTER 2

Principles of Power

2.1 Background on Power

Power consumption in CMOS circuits consists of three major components, assuming that the DC power consumption (static) is zero. This is almost always the case with CMOS logic families. A general expression for power consumption is shown in Figure 2.1. The parameter α is a switching activity parameter that can range from 0 to 1 depending on how often the output of the gate switches logic levels and is a function of both the path and type of logic gate itself. The first (and usually dominant) source of power consumption is due to dynamic switching power needed to drive the capacitive loads on gates. The second component is due to the short circuit current (if any) which occurs for some amount of time during a switching event. ("*Active energy or power*" as it is referred

$$P \sim \alpha \cdot \left(C_L \cdot V_{swing} + \overline{I_{SC}} \cdot \Delta t_{SC}\right) \cdot V_{DD} \cdot f + \left(I_{DC} + I_{Leak}\right) \cdot V_{DD}$$

- α - switching probability
- C_L – load capacitance
- V_{swing} – voltage swing
- f - frequency
- I_{SC} – mean value of switching transient current
- Δt_{SC} – short current time
- I_{DC} – static current
- I_{leak} – leakage current

Dominant: $P \sim \alpha \cdot C_L \cdot V_{swing} \cdot V_{DD} \cdot f$

Figure 2.1 Power Consumption in CMOS Circuits [Nikolic01].

to hereafter in this document is comprised of these two components.) The third component is static power consumption and is generally zero for most logic families. Finally, a leakage power component makes up the fourth component. This leakage power is rapidly gaining the attention of VLSI designers today as discussed earlier. It is attributed to leakage current, which is related to threshold voltage as follows [Kuroda98]:

$$I_{LEAK} = \frac{I_0}{W_0} W \cdot 10^{\frac{V_T}{S}} . \tag{2.1}$$

W is the channel width, and S is the subthreshold slope. The typical value of S is 0.1V/decade, which reflects an order of magnitude increase in leakage with a 0.1V drop in threshold voltage. In a typical deep submicron technologies, I_0 and W_0 are 1μA and 10μm, respectively.

In order to find the tradeoff between power and delay in CMOS circuits the delay of a gate must be characterized accurately with varying threshold voltage and supply voltage. An elegant model for this characterization is the Alpha-power law, first presented in [Sakurai90]. The basis of this model lies with the fact that the current in a short-channel MOSFET does not really increase with the square of the overdrive voltage but rather increases with some other power α, which is generally between 1 and 2. Using this fact, the drain current can be expressed as [Sakurai90]:

$$I_D = \begin{cases} 0, & (V_{GS} \le V_{TH} : cutoff\ region) \\ (I'_{D0} / V'_{D0}) V_{DS}, & (V_{DS} < V'_{D0} : triode\ region) \\ I'_{D0}, & (V_{DS} \ge V'_{D0} : pentode\ region) \end{cases}, \tag{2.2}$$

where I'_{D0} is given by:

$$I'_{D0} = I_{D0} \left(\frac{V_{GS} - V_{TH}}{V_{DD} - V_{TH}} \right)^{\alpha} \left(= \frac{W}{L_{eff}} P_C (V_{GS} - V_{TH})^{\alpha} \right) \tag{2.3}$$

and V'_{DO} is:

$$V'_{D0} = V_{D0} \left(\frac{V_{GS} - V_{TH}}{V_{DD} - V_{TH}} \right)^{\alpha/2} \left(= P_V (V_{GS} - V_{TH})^{\alpha/2} \right). \tag{2.4}$$

Using this expression for drain current, the delay of a CMOS inverter can be derived as [Sakurai90]:

$$t_{pHL},t_{pLH}=\left(\frac{1}{2}-\frac{1-v_T}{1+\alpha}\right)t_T+\frac{C_L V_{DD}}{2I_{D0}},\quad V_T=\frac{V_{TH}}{V_{DD}}. \tag{2.5}$$

In addition to these two expressions, the alpha-power law can be used to determine the delay as a function input slope and source/drain resistance. These results are derived in [Sakurai90].

2.2 Multiple Threshold Voltages

As technology scaling has lowered the supply voltage, threshold voltages have been scaled down to maintain performance. Since subthreshold leakage currents increase exponentially as threshold voltage is reduced, the increase in leakage power is significant. Currently, subthreshold leakage is not the dominant source of power consumption in CMOS circuits, but it may soon outpace dynamic power dissipation [Kao00].

In order to reduce leakage current without compromising performance, multiple threshold voltages are used. Typically, the process cost of each additional threshold voltage is one mask step each for nMOS and pMOS devices. Most sub-0.25μm CMOS technologies do offer two types of nMOS and pMOS devices with thresholds differing by about 100mV, adjusted by channel doping. This 100mV difference corresponds 10x difference in leakage. The delay increase associated with the higher threshold is dependent on both process parameters and the supply voltage.

The use of multiple threshold voltages has been discussed in various ways, such as the use of low-V_T devices only on the critical paths [Kato00]. Another approach is to use low-V_T devices for CMOS static circuits, and high-V_T devices for dynamic circuits, which are noise margin sensitive [Thompson97]. In [Kato00], a scheme termed "Random Modulation" is proposed in which threshold voltage is assigned on a cell-by-cell basis with the objective of minimizing the number of low-V_T cells without degrading performance. This technique allows for further power reduction than the case of "binary modulation," in which threshold voltages are assigned on a per-path basis (see Figure 2.2).

The low threshold voltage should be chosen in order to meet the targeted frequency, while the high threshold voltage should be chosen in order to minimize leakage. If the high-V_T is too high, only a small amount of high-V_T cells will be able to be assigned while maintaining performance, minimizing the impact of these cells. However, if the high-V_T is set too close to the low-V_T, the leakage current of the high-V_T cells will be too high for optimal reduction [Kato00].

In order to assign cells, [Kato00] presents an algorithm. The purpose of the algorithm is to minimize the number of low-V_T cells, which can be somewhat complicated due to cells belonging to more than one path. The first step is to change all cells to high-V_T and then perform delay and slack calculations. Starting with the path with the least violated slack, for each cell in the path, an evaluation value is calculated by use of an evaluation function:

$$EF(cell) = \frac{t_{pd}(cell)}{I_{sub}(cell)} \times N_{vp}(cell). \tag{2.6}$$

This function enables the selection of a cell which will result in a large reduction of delay, possibly in multiple violated paths, and a small increase in leakage current. The cell with the highest evaluation value is then replaced with a low-V_T cell. The delay and slack is then recalculated. While the path slack is still negative, more cells are replaced. This entire process continues until all the negative slack paths have been resolved or all

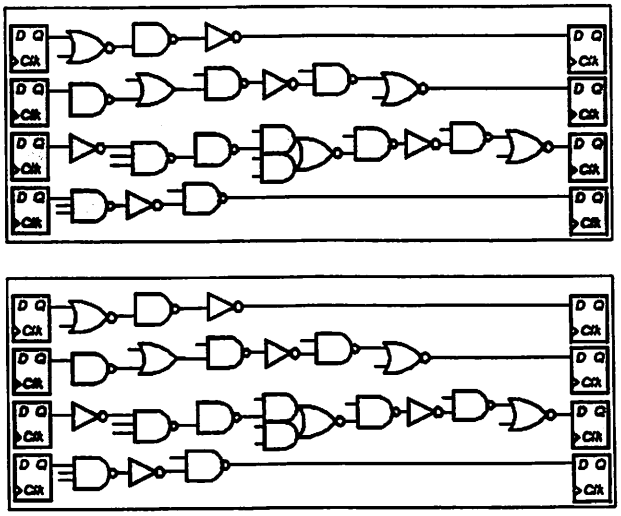


Figure 2.2 Binary modulation (top) and random modulation, low-V_T shaded [Kato00].

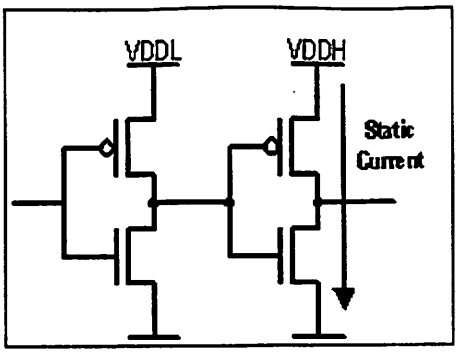


Figure 2.3 Static current from V_{DDL} to V_{DDH}.

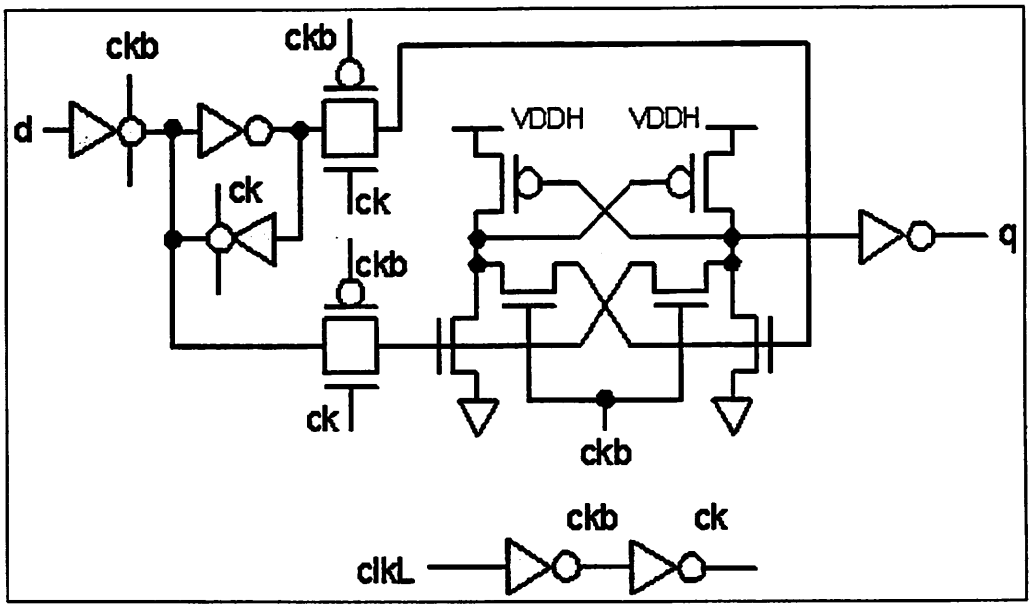


Figure 2.4 Level-converting flip-flop.

there are no more cells to be replaced.

2.3 Multiple Supply Voltages

As shown in 2.1 (page 5), dynamic power is proportional to the product of the supply voltage and the voltage swing. In CMOS circuits, this reduces to V_{DD}^2 in most cases. Therefore, a reduction in supply voltage causes a quadratic decrease in active power, along with a linear decrease in leakage power. However, a supply voltage reduction causes performance degradation, which is dependent on threshold voltage and process parameters. In order to preserve performance, while also reducing power, a dual-V_{DD} approach can be used. Gates off the critical path are run at a reduced supply voltage, V_{DDL}, while those on the critical path are run at V_{DDH}.

Some important issues must be considered with the use of multiple supply voltages. The output of V_{DDL} gates cannot be fed directly to V_{DDH} gates because the output of a V_{DDL} gate can never be raised higher than V_{DDL}. Therefore, if connected to a V_{DDH} circuit, static current will flow due to the PMOS in the V_{DDH} circuit never being completely cut-off, see Figure 2.3 [Usami95].

In order to block this current, level converters can be inserted between V_{DDL} and V_{DDH} circuits. Level converters add additional area and power overhead, and therefore should be minimized. One approach is to embed the level shifting function within a flip-flop circuit (LCFF), see Figure 2.4. In [Usami00], it was reported that this results in the power of the flip-flop being less than that of a V_{DDH} flip-flop, while increasing delay only slightly. Asynchronous level converters can also be used, although they carry an increased area penalty compared to LCFF.

Layout is another important issue when dealing with multiple supply voltages. V_{DDL} and V_{DDH} cells must be separated due to different N-well voltages. Generally a row-by-row separation, Figure 2.5, is used [Usami95][Kuroda98] due to high-performance and applicability to both standard-cell and gate-arrays. Since cells are now assigned to layout rows based on cell supply voltage, algorithms must be used to determine optimal row voltage assignment. Balancing of the rows is essential to avoid increased area penalties, although this is not always possible. Additional area penalties are incurred by increased routing. Additional bypass capacitance may also be required with a multiple supply

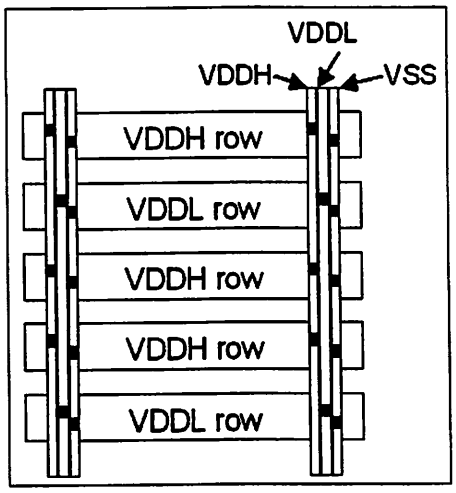


Figure 2.5 Dual-supply layout, row-by-row supply voltage separation.

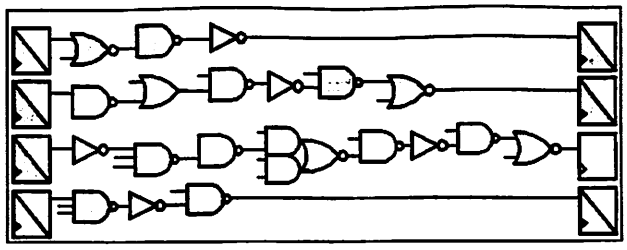


Figure 2.6 Clustered voltage scaling structure, low-V_{DD} shaded.

design.

Clustered Voltage Scaling (CVS) is one structure proposed to implement dual-V_{DD} design. With CVS, all paths are structured as follows: primary inputs → V_{DDH} cells → V_{DDL} cells → level converters → primary outputs, as shown in Figure 2.6. This formation leads to clusters of V_{DDH} cells and V_{DDL} cells. By inserting level converters only at the end of a path, the number needed is minimized. [Usami00]

If the primary outputs are latched at the end of the path, a FFLC can be used. To assign the V_{DDL} cells, a depth-first-search algorithm is used from the primary outputs to the primary inputs. As each cell is visited, an attempt is made to replace it with a V_{DDL} cell. If it can be replaced, the algorithm continues, otherwise the traversal is stopped. Dealing with multiple fanouts can be tricky. In order to replace a cell, all of the cells in the fanout of that cell must also be replaced with V_{DDL} cells [Usami95].

Extended CVS (ECVS) differs from CVS in that it allows placement of level converters even between logic gates. This can be useful in the case where a gate has multiple inputs and only one is on the critical path. Before a level converter is inserted between logic gates, it is checked to see if the insertion does reduce power [Usami00]. ECVS shows only marginal improvement and the need for asynchronous level converters makes it unattractive. In [Usami97], an automated CAD tool was used to implement ECVS. Asynchronous level converters were inserted only where advantageous, for instance, at a gate with multiple inputs where only one is on the critical path, and where the power overhead of the level converter did not eclipse the power savings.

CHAPTER 3

Theoretical Study

3.1 Rules of Thumb Derived in [Hamada01]

In [Hamada01], a theoretical approach was used to formulate rules of thumb for optimal V_{DD}'s, V_T's and transistor widths for maximum power reduction. First considered was the use of multiple power supplies.

3.1.1 Rule of Thumb for Multiple Supplies

With multiple power supplies, $V_1 > V_2 > \ldots > V_n$, switching power dissipation can be expressed as:

$$P_n = f \cdot \left\{ \left(C_{TOT} - \sum_{i=2}^{n} C_i \right) \cdot V_1^2 + \sum_{i=2}^{n} C_i \cdot V_i^2 \right\}, \tag{3.1}$$

where C_{TOT} is the total capacitance of the circuit, and C_i is the capacitance that will operate under V_i. For clarity, consider the case of two power supplies:

$$P_2 = f \cdot \left[(C_{TOT} - C_2) \cdot V_1^2 + C_2 \cdot V_2^2 \right]. \tag{3.2}$$

The ratio of power dissipation, comparing multiple supplies to a single power supply, is given by:

$$R_{VDD} \equiv \frac{P_n}{P_1} = 1 - \sum_{i=2}^{n} \left[\left(\frac{C_i}{C_{TOT}} \right) \cdot \left\{ 1 - \left(\frac{V_i}{V_1} \right)^2 \right\} \right]. \tag{3.3}$$

Limiting to the case of dual-supply, this equation reduces to:

$$R_{VDD} \equiv \frac{P_2}{P_1} = 1 - \left[\left(\frac{C_2}{C_{TOT}} \right) \cdot \left\{ 1 - \left(\frac{V_2}{V_1} \right)^2 \right\} \right]. \tag{3.4}$$

By profiling a typical design, the authors determined that delay and capacitance are roughly in proportion to one another, i.e. this assumes the longest path has the most capacitance. This enabled the following substitution for the ratio of capacitances:

$$\frac{C_i}{C_{TOT}} = \frac{\int_0^1 p(t) \cdot t_i \cdot dt}{\int_0^1 p(t) \cdot t \cdot dt}. \tag{3.5}$$

In this equation, *p(t)* represents the normalized path-delay distribution and t_i is the total delay of circuits that will operate at V_i. The calculation of t_i is as follows:

$$\begin{aligned} t_i &= \frac{t_{i.0}}{t_{i.0} - t_{(i+1).0}} \left(t - t_{(i+1).0} \right) \qquad (t_{(i+1).0} \le t \le t_{i.0}) \\ t_i &= \frac{t_{i.0}}{t_{(i-1).0} - t_{i.0}} \left(t_{(i-1).0} - t \right) \qquad (t_{i.0} \le t \le t_{(i-1).0}), \end{aligned} \tag{3.6}$$

where $t_{i.0}$ is derived from the alpha-power law model as:

$$\begin{aligned} t_{i.0} &= \left(\frac{V_1}{V_i} \right) \cdot \left(\frac{V_i - V_T}{V_1 - V_T} \right)^{\alpha}. \\ t_{n+1.0} &= 0 \end{aligned} \tag{3.7}$$

For example, for the case of two supplies, $t_1 = 1$ and let $t_2 = 0.5$ (i.e. a path with a normalized delay of 0.5 has a normalized delay of 1 when placed in the second supply), if you have a path with delay, $t = 0.75$:

$$t_1 = \frac{1}{1-0.5}(0.75 - 0.5) = 0.5; \quad t_2 = \frac{0.5}{1-0.5}(1 - 0.75) = 0.25. \tag{3.8}$$

In this example path, 0.5 of a normalized capacitance value would be switched at V_1 and 0.25 would be switched at V_2.

Again limiting to the case of dual-supply, substituting equation 3.6 into equation 3.5 gives:

$$\frac{C_2}{C_{TOT}} = \frac{\int_0^{t_{2.0}} p(t)\cdot t\cdot dt + \int_{t_{2.0}}^{1} p(t)\cdot \frac{t_{2.0}}{1-t_{2.0}}(1-t)\cdot dt}{\int_0^1 p(t)\cdot t\cdot dt}, \tag{3.9}$$

The total capacitance of all paths with delay $\leq t_2$ is switched at V_2, while this is true for only a portion of the capacitance of paths with delay $> t_2$.

Using equations 3.3 and 3.5 through 3.7, the power dissipation ratio, R_{VDD}, can be calculated for a given $p(t)$, V_1, V_i, and V_T. The authors used these equations to develop a rule of thumb for optimum supply voltages, shown in Figure 3.1, with lambda-shaped $p(t)$ as a basis. When compared with simulation, the rule of thumb showed power reduction within 1% from the absolute minimum.

$$For\{V_1,V_2\}: \frac{V_2}{V_1} = 0.5 + 0.5\frac{V_T}{V_1}$$

$$For\{V_1,V_2,V_3\}: \frac{V_2}{V_1} = \frac{V_3}{V_2} = 0.6 + 0.4\frac{V_T}{V_1}$$

$$For\{V_1,V_2,V_3,V_4\}: \frac{V_2}{V_1} = \frac{V_3}{V_2} = \frac{V_4}{V_3} = 0.7 + 0.3\frac{V_T}{V_1}$$

Figure 3.1 Rule of thumb for optimum multiple supply voltages from [Hamada01].

3.1.2 Rule of Thumb for Multiple Thresholds

Leakage current in a chip with multiple threshold voltages, $V_{T.1} < V_{T.2} < \ldots < V_{T.n}$, is given by:

$$I_n = \left(\frac{I_0}{W_0}\right)\cdot\left\{\left(W_{TOT} - \sum_{i=2}^{n} W_i\right)\cdot 10^{\frac{-V_{T.1}}{S}} + \sum_{i=2}^{n} W_i \cdot 10^{\frac{-V_{T.i}}{S}}\right\}, \tag{3.10}$$

where W_i represents the total gate width of pMOS and nMOS devices with threshold voltage V_{Ti} and whose source is connected to V_{DD} and V_{SS}. The ratio of leakage current, comparing multiple thresholds to a single threshold, is given by:

$$R_{VT} \equiv \frac{I_n}{I_1} = 1 - \sum_{i=2}^{n}\left[\left(\frac{W_i}{W_1}\right)\cdot\left\{1 - 10^{\frac{V_{T.i}-V_{T.1}}{S}}\right\}\right]. \tag{3.11}$$

Design profiling showed delay is in proportion to the sum of gate width of transistors connected to V_{DD} or V_{SS}. Therefore the ration of W_i to W_1 can be calculated in the same manner C_i/C_1 was calculated in 3.1.1 (page 12).

The authors used these equations to formulate a rule of thumb for optimum threshold voltages, Figure 3.2.

$$
\begin{aligned}
&For\{V_{T1},V_{T2}\}: V_{T2} = 0.10V_{DD} + V_{T1} \\
&For\{V_{T1},V_{T2},V_{T3}\}: V_{T2} = 0.06V_{DD} + V_{T1} \\
&\longrightarrow V_{T3} = 0.07V_{DD} + V_{T2} \\
&For\{V_{T1},V_{T2},V_{T3},V_{T4}\}: V_{T2} = 0.04V_{DD} + V_{T1} \\
&\longrightarrow V_{T3} = 0.05V_{DD} + V_{T2} \\
&\longrightarrow V_{T3} = 0.06V_{DD} + V_{T3}
\end{aligned}
$$

Figure 3.2 Rule of thumb for optimum multiple threshold voltages from [Hamada01].

3.1.3 Rule of Thumb for Transistor Sizing

Switching power in a chip employing multiple transistor widths, $W_1 > W_2 > \ldots > W_n$, can be expressed as:

$$P_n = f \cdot \left\{ \left(C_{MOS.1} - \sum_{i=2}^{n} C_{MOS.i} \right) + \sum_{i=2}^{n} \left(C_{MOS.i} \cdot \frac{W_i}{W_1} \right) + C_{INT} \right\} \cdot V^2 , \qquad (3.12)$$

with $C_{MOS.i}$ representing the total gate and diffusion capacitance of transistors with width scaled to W_i and C_{INT} representing the total interconnect capacitance. The ratio of power dissipation, comparing multiple transistor widths to a single width, is given by:

$$R_W \equiv \frac{P_n}{P_1} = 1 - \frac{1}{m+1} \sum_{i=2}^{n} \left(\frac{C_{MOS.i}}{C_{MOS.1}} \right) \cdot \left(1 - \frac{W_i}{W_1} \right), \qquad (3.13)$$

where m is $C_{INT}/C_{MOS.1}$. The same method as in 3.1.1 (page 12) can be followed to calculate the ratio of W_i to W_1. The authors calculated a rule of thumb for optimum transistor widths, Figure 3.3, from these equations.

$$For\{W_1, W_2\}: W_2 = \frac{1}{2}W_1$$

$$For\{W_1, W_2, W_3\}: W_2 = \frac{2}{3}W_1, W_3 = \frac{1}{3}W_1$$

$$For\{W_1, W_2, W_3, W_4\}: W_2 = \frac{3}{4}W_1, W_3 = \frac{1}{2}W_1, W_3 = \frac{1}{4}W_1$$

Figure 3.3 Rule of thumb for optimum multiple transistor widths from [Hamada01].

3.2 Limitations on Rules of Thumb

[Hamada01] presents a comprehensive solution to the problem of determining supply voltages, threshold voltages, and transistor sizes for power reduction in multi-supply, multi-threshold and multi-size designs, respectively. However, the case where these methods of power reduction are combined was not addressed. There is no simple method to combine these techniques into a single rule of thumb. In [Hamda01], power was reduced to switching power for the case of multiple supply voltages and to leakage power for multiple threshold voltages. To combine these two rules of thumb, a more complete power model would be required, such as Figure 2.1, page 5.

The effect of level-conversion penalties on the multi-supply was not considered. With a traditional level converting flip-flop design (see Figure 2.4, page 9), the delay penalty for level conversion can be quite large. This can have an effect on the fraction of the total capacitance that can be placed in the lower V_{DD}. However, innovative level-converting flip-flop (LCFF) designs shown in [Ishihara02] demonstrate that delay penalties can be adjusted so that the delay scales similarly to gate delays. An added concern is the additional power penalty of the LCFF.

CHAPTER 4

Test Environment

4.1 Overview of Design Framework

Early on in this project, it was determined that adequate CAD software support was not available to easily examine the effects of multiple threshold voltages, sizing, and especially multiple supply voltages. In order to facilitate this exploration by avoiding lengthy simulation cycles, a design framework was constructed using MS Excel software and Spectre simulations. First, basic gate models were derived from simulation. A structure was then built in a MS Excel spreadsheet in which gates can be combined to form paths. The delay, active energy, and leakage power for each path is calculated through the use of the gate models and the sizing of the gates.

4.2 Preliminary Investigation

Before the main body of this work was undertaken, a preliminary investigation was conducted to determine the direction of the research. The central focus of this study was dual-threshold and multiple-supply design in 0.18μm, with some consideration of multiple-length design. Sizing was not considered. This inquiry made use of a rough version of the design framework described above. The gate models used were simplified. Figures for a gate's power/energy consumption and delay were not calculated based on the actual load capacitance, but rather on a preset value. (See Appendix A for a detailed listing of the gate models.) Neglecting the penalties associated with the use of level-converting flip-flops further simplified the study.

Table 4.1 Comparison of power reduction techniques.

Technique	Active Energy [pJ]	Normalized Active Energy	Leakage Power [uW]	Normalized Leakage Power
Baseline	43.78	1.0	4.95	1.0
Dual-V_T, Single-V_{DD}	41.94	0.96	0.62	0.13
Single-V_T, Dual-V_{DD}	27.11	0.62	2.78	0.56
Single-V_T, Triple-V_{DD}	23.58	0.54	2.29	0.46
Dual-V_T, Dual-V_{DD}	26.63	0.61	1.91	0.39

Table 4.1 summarizes the power reduction achieved by applying combinations of dual-threshold and multi-supply techniques to the initial design. The supply voltages used were determined using the rule of thumb developed in [Hamada01]. The nominal supply voltage was 1.8V. 1.8V and 1.1V were used for dual-supply and 1.8V, 1.24V and 0.86V were used for triple-supply.

4.2.1 Impact of Triple-Supply Voltages

The preliminary investigation showed active power reduction of 38% when using dual-supply voltage and an increase to 46% reduction with triple-supply. The improvement between dual- and triple-supply is minimal. The use of dual-supply may prove a useful technique with further study, while it seems somewhat unlikely that three or more supply voltages will be used heavily due to the added design complexity and overhead.

4.2.2 Effect of Channel Length Variation

The use of multiple channel lengths was considered. Simulations were done for each basic gate at three different channel lengths, 0.18μm, 0.19μm and 0.20μm, using high-speed transistors, and also at the nominal 0.18μm channel length with low-leakage transistors. Figure 4.1 details the results averaged across gates. Modifying transistor channel lengths is comparable to using a 2nd threshold in terms of the leakage reduction vs. delay increase tradeoff. However, increasing channel length has a negative side effect of increasing active energy due to the increase in gate capacitance. Using high threshold

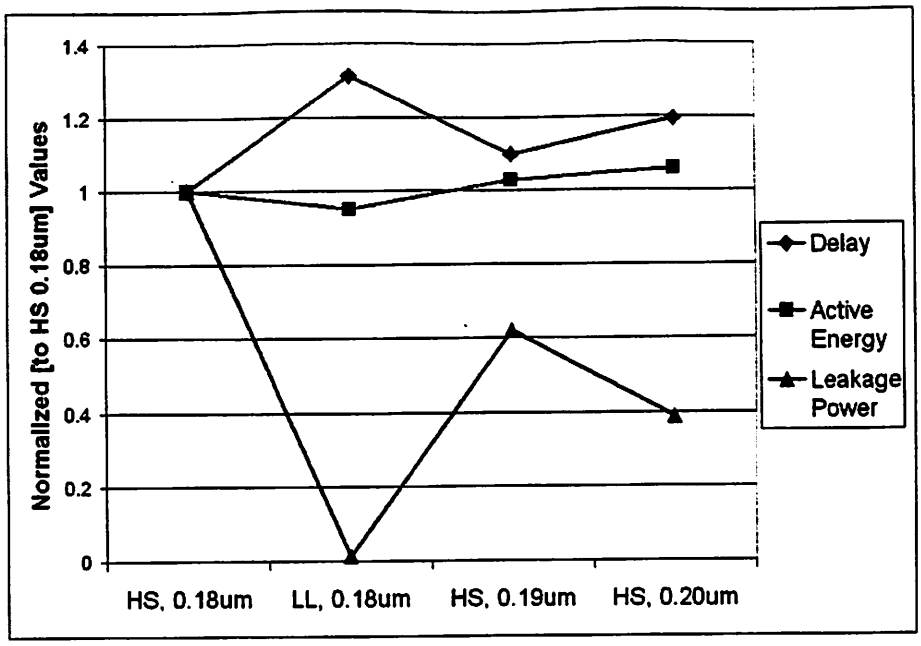


Figure 4.1 Effect of channel length modulation compared with a 2nd threshold voltage.

devices actually causes a small reduction in active power due to reduced gate channel capacitances in the off state and a small reduction in signal swings (V_{DD}-V_T). Therefore, using channel length modification to control leakage does not seem to be a practical solution. One exception may be cases in which finer control of the delay increase is required.

4.2.3 Conclusions from Preliminary Investigation

The use of multiple channel lengths was shown to be unfavorable from a cost-benefit standpoint when compared with the use of multiple-thresholds. Both the dual-V_{DD} and dual-V_T techniques were shown to have significant power savings, separately and in conjunction. However, not only was the FFLC penalty neglected, another aspect of circuit design, sizing, was left out altogether. From the preliminary results, it was decided to focus on dual-V_{DD}, dual-V_T and sizing in the next phase of experimentation.

4.3 Power and Delay Models

This work is based on a linear delay model, where the delay is expressed as a linear function of the load capacitance. In this early evaluation, the delay dependence on input

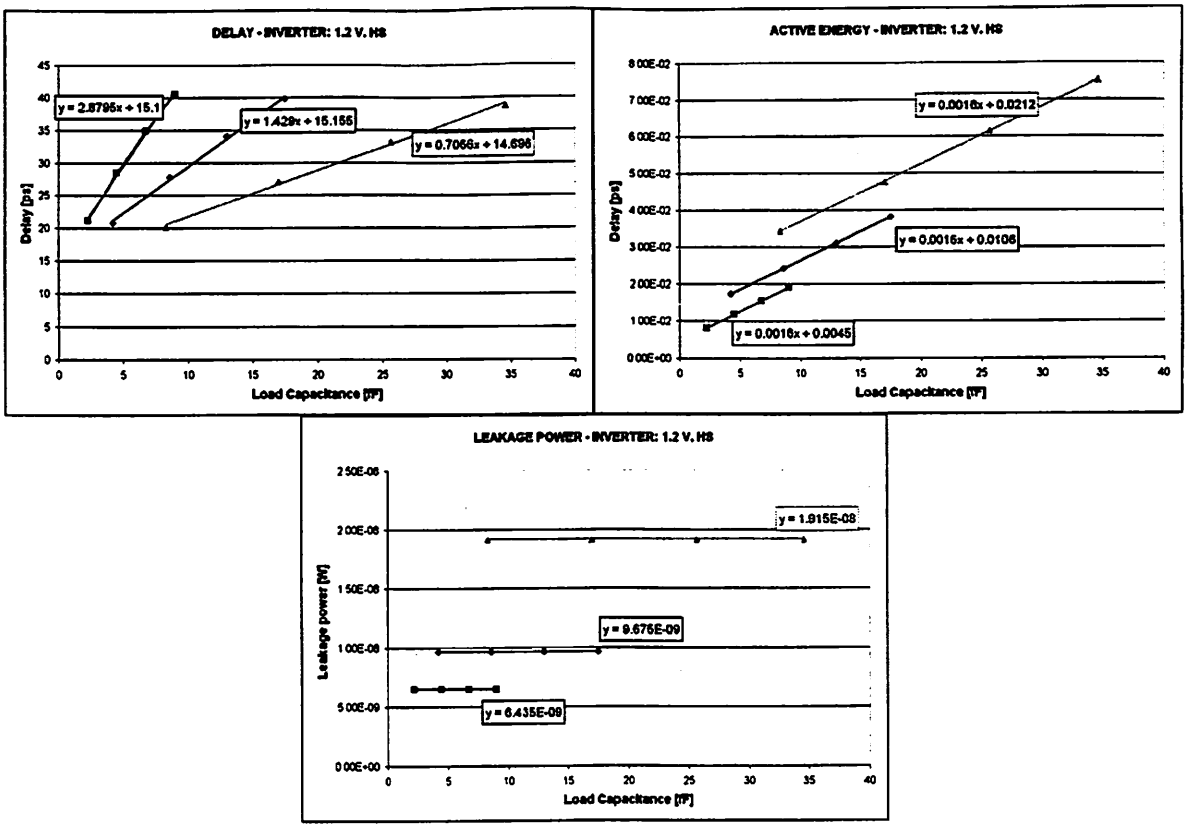


Figure 4.2a-c Effect of sizing on delay (a), active energy (b) and leakage power (c).

slope is ignored. A simplified library is based on 7 logic gates, each of them designed with multiple sizes. Using Cadence™ Composer and Spectre, these gates were simulated using a 0.13μm process The technology provides two threshold voltages were available, high speed (HS), or low V_T, and low-leakage (LL), or high V_T, with a spread of approximately 100mV. The baseline supply voltage was 1.2V. Two values were evaluated for the second supply voltage, V_{DDL}: 0.8 V, chosen according to [Hamada01], and 1.0V, as an alternate choice.

Each gate/supply/threshold combination was simulated to determine active energy, leakage power and delay for several different load capacitances. To complete the models, delay and active energy were plotted against the load capacitance. Linear extrapolation was used to determine the slope and y-intercept values for both active energy and delay. The gate models, consisting of leakage power values, y-intercept and slope values for

both delay and active energy, and input capacitance, for each combination of supply and threshold are listed in Appendix B.

In order to determine how the models should be adjusted for transistor sizing, simulations were performed for 2x and 4x inverters. Figure 4.2a-c demonstrates the effect of sizing on delay, active energy and leakage power.

4.4 Level-Converting Flip-Flops

As shown earlier, it is necessary to insert a level-converter between V_{DDL} gates and V_{DDH} gates. For purposes of this project it was decided to use a Clustered Voltage Scaling (CVS) scheme for insertion of low supply voltage gates. Since all low voltage gates in a path are clustered together at the end of path, the level conversion function can be combined with the flip-flop (FF).

4.4.1 Traditional Level-Converting Flip-Flop

The traditional level converting flip-flop (LCFF) design is shown in Figure 2.4 (page 9). Simulation revealed large penalties for using the traditional LCFF design when compared with a basic flip-flop (Table 4.2). This design is unsuitable for a dual-V_{DD} design. The large delay penalties would severely limit the number of gates that could be placed at the lower supply voltage.

Table 4.2 Delay/energy of flip-flop and traditional LCFF.

V_{DDL} [V]	Setup Time [ps]	Clk-to-Q Delay [ps]	Total Delay [ps]	Normalized Energy
Baseline (FF)	31	140.5	171.5	1.0
0.8V	135	478.5	613.5	1.45
1.0V	87.5	197.5	285	1.10

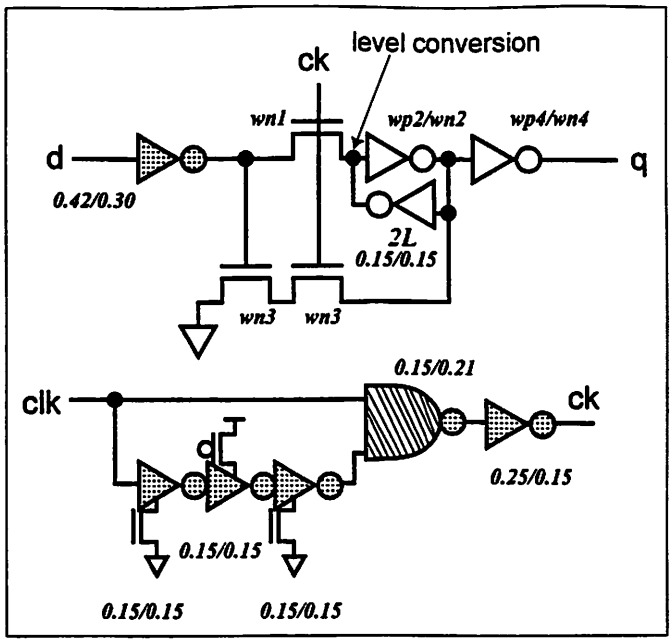


Figure 4.3 Schematic of improved LCFF.

Table 4.3 Delay/energy of flip-flop and improved LCFF.

V_{DDL} [V]	Setup Time [ps]	Clk-to-Q Delay [ps]	Total Delay [ps]	Normalized Energy
Baseline (FF)	14.5	130	144.5	1.0
0.8V	37.5	143	180.5	2.8
1.0V	6.5	130	136.5	1.29

4.4.2 New LCFF Design

In [Ishihara02], innovative LCFF designs were presented. The design most suited to this project, see Figure 4.3, was chosen as an alternative to the traditional LCFF discussed in 4.4.1 (page 21). This design effectively eliminates the delay penalty at V_{DDL}=1.0V and drastically reduces it at V_{DDL}=0.8V. These simulation results are shown in Table 4.3. The downside of this design is its higher power consumption. This power penalty is offset in a dual-V_{DD} design by the increase in the number of gates that can be placed in V_{DDL} with this LCFF. Another disadvantage is a long hold time, but this should not be a problem in dual supply designs since V_{DDL} is applied to slow down short paths.

4.5 Baseline Design

A baseline design was formed by randomly stringing together combination of gates, up to 12 gates per path, to form 500 paths. Initially, all gates were minimum size. Sizing was then used to bring each path to its maximum speed to approximate a synthesized design. Loading for each path was set to 50x the input capacitance of a flip-flop, or 175fF. A large capacitance was chosen to approximate driving a long bus. Figure 4.4 shows the effect of creating the baseline design through sizing on the path-delay distribution of the initial unsized design. Active energy, leakage power and maximum delay numbers for the two designs are listed in Table 4.1.

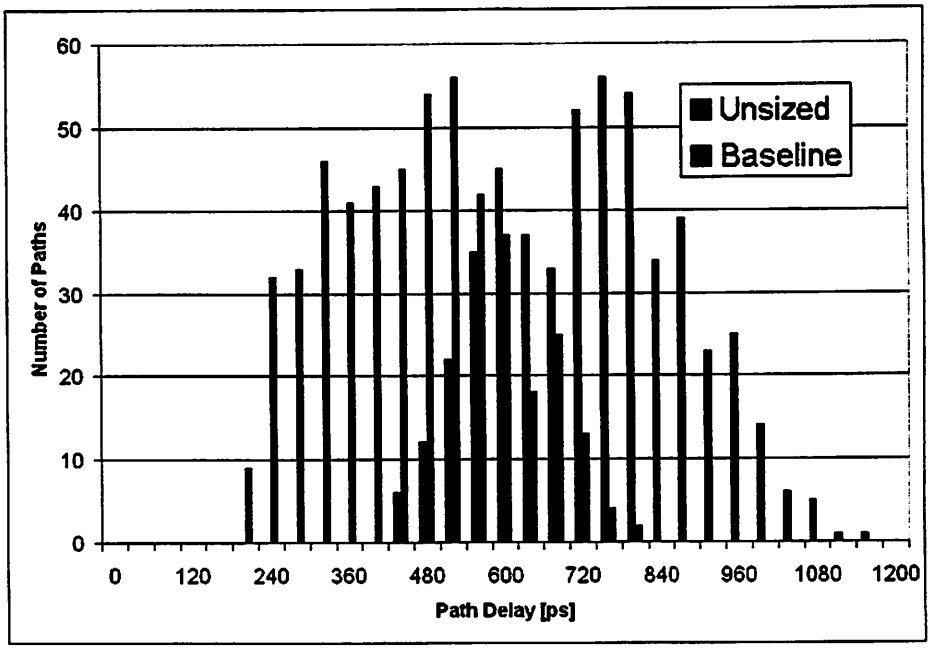


Figure 4.4 Path-delay histograms of baseline and initial unsized designs.

Table 4.4 Summary of baseline and unsized designs.

Design	Active Energy [pJ]	Leakage Power [μW]	Maximum Delay [ps]
Initial unsized design	214.9	45.86	1142.6
Baseline design	293.3	76.80	721.2

CHAPTER 5

Analysis of Results

Dual-threshold, dual-V_{DD} and sizing techniques were applied to the baseline design separately and in conjunction in order to evaluate the effectiveness of these techniques. The performance of the baseline was preserved in all cases. Complete numerical results are listed in Appendix C.

The baseline design represents a general logic block with a lambda-shaped path-delay distribution (see Figure 4.4, page 23). From [Hamda01], lambda-shaped distribution is consistent with after-layout static timing analysis. The framework developed in 4.1, page 17, prevents the modeling of diverging and reconvergent paths. Therefore, each gate is limited to a fanout of 1 for now.

5.1 Sizing

Here, gates of the critical paths were downsized where possible. Figure 5.1 details the effect of sizing on the path-delay distribution, while Figure 5.2 shows the normalized active energy and leakage power for the sized design against the baseline design. In essence, most gates off the critical path are minimum sized.

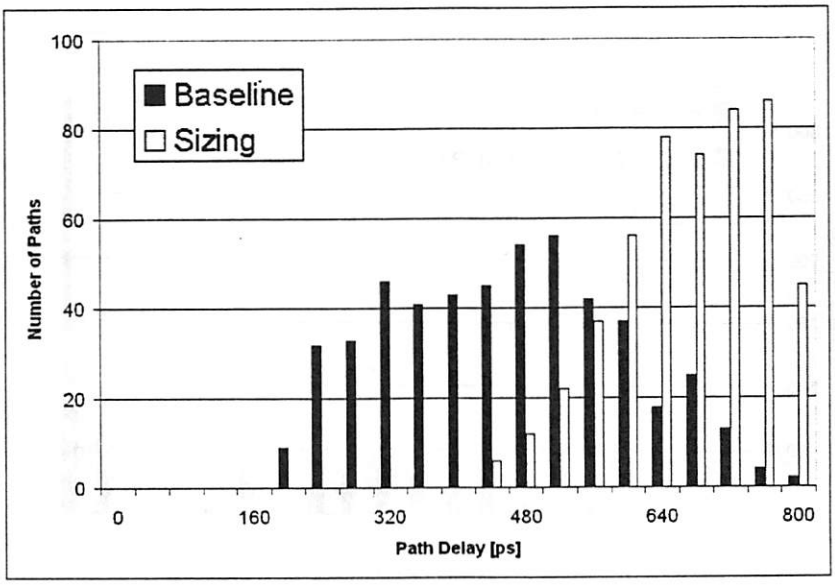


Figure 5.1 Effect of sizing on path/delay distribution.

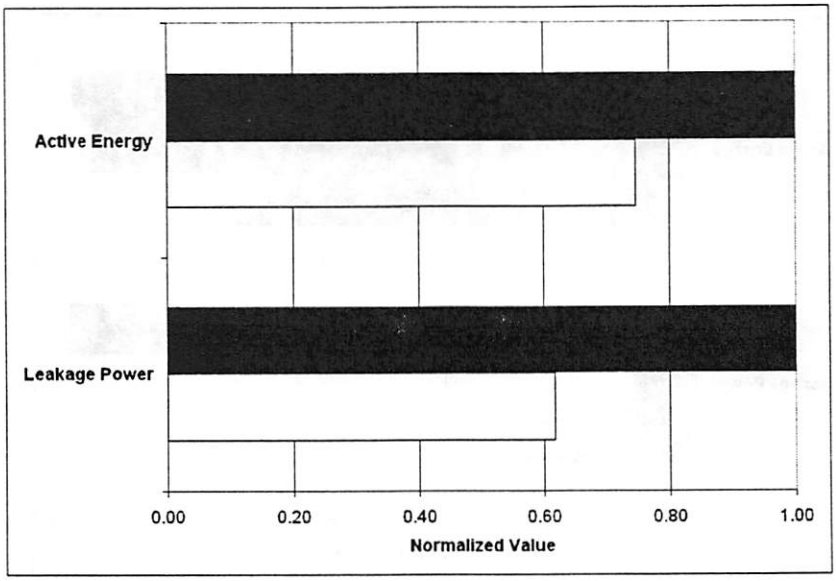


Figure 5.2 Effect of sizing on power/energy.

5.2 Dual Threshold Voltages

The dual-threshold technique was applied to both the baseline design and the sized design. Cell assignment (to either LL or HS) was done on a cell-by-cell basis, with the goal being maximum reduction of leakage energy for each path while still meeting timing goals. By replacing high-speed cells with low leakage cells, leakage power was reduced substantially in both the baseline design and the sized design. See Figure 5.3 for the path-delay distribution, and Figure 5.4 for the normalized power/energy numbers. Since

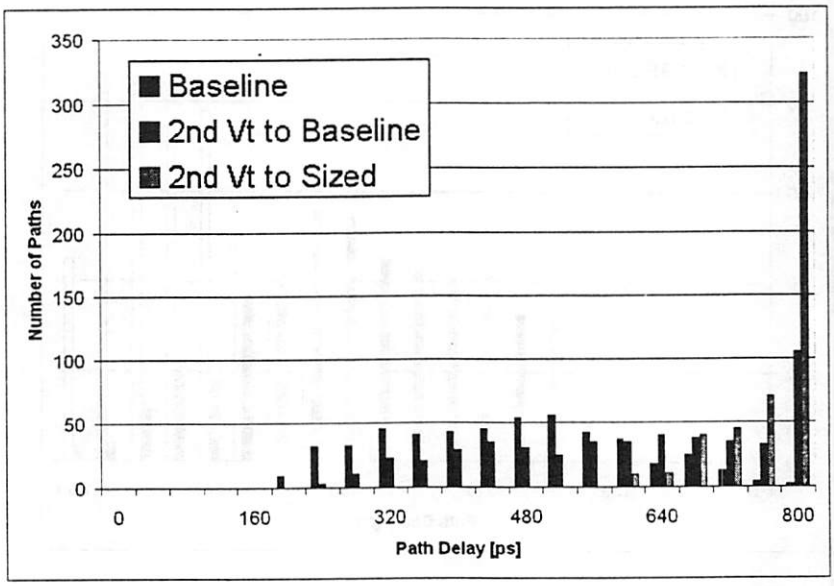


Figure 5.3 Effect of dual-threshold on path/delay distribution.

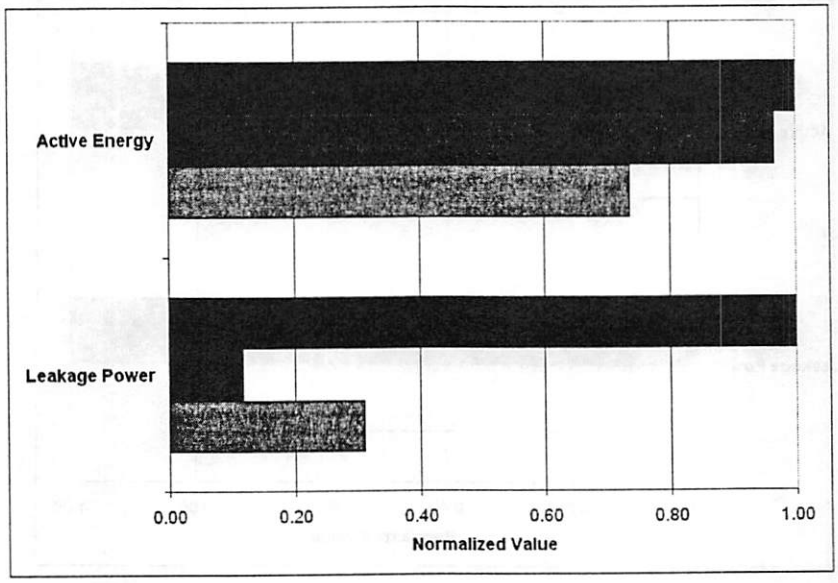


Figure 5.4 Effect of dual-threshold on power/energy.

leakage power was reduced more substantially when dual-threshold was applied to the baseline design as opposed to when it was applied to the sized design, it can be concluded that dual-threshold is more powerful than sizing for leakage power control, as expected. There is a small active power reduction with dual-threshold.

5.3 Dual Supply Voltages

When applying the dual-supply technique, clustered voltage scaling was used (see 4.4, page 21). The dual-supply technique was applied to both the baseline design and the sized design.

5.3.1 Low Supply Voltage of 0.8V

The path-delay distributions for the dual-supply design are shown in Figure 5.5, and the normalized power/energy in Figure 5.6. It is unclear here if sizing or dual-supply is more effective for active energy reduction. It is apparent, however, that dual-supply is a better choice than sizing for decreasing leakage power.

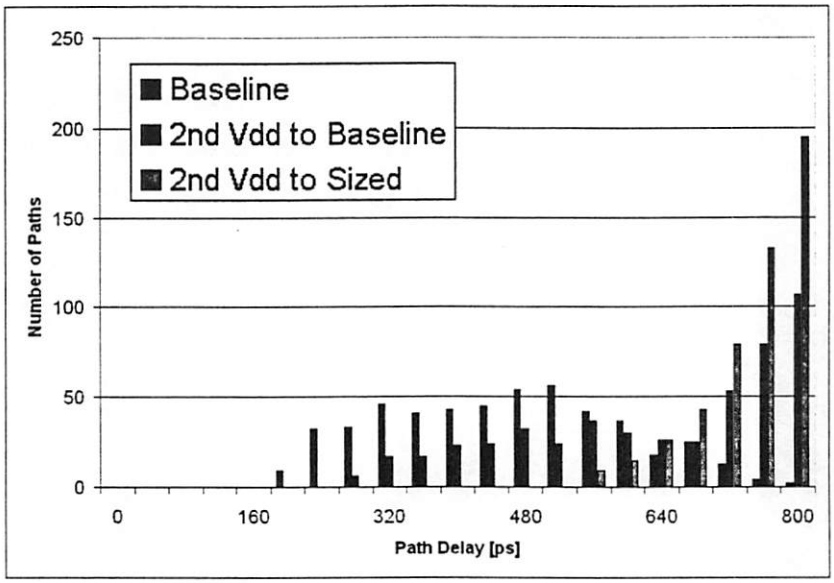


Figure 5.5 Effect of dual-supply (V_{DDL} = 0.8V) on path-delay distribution.

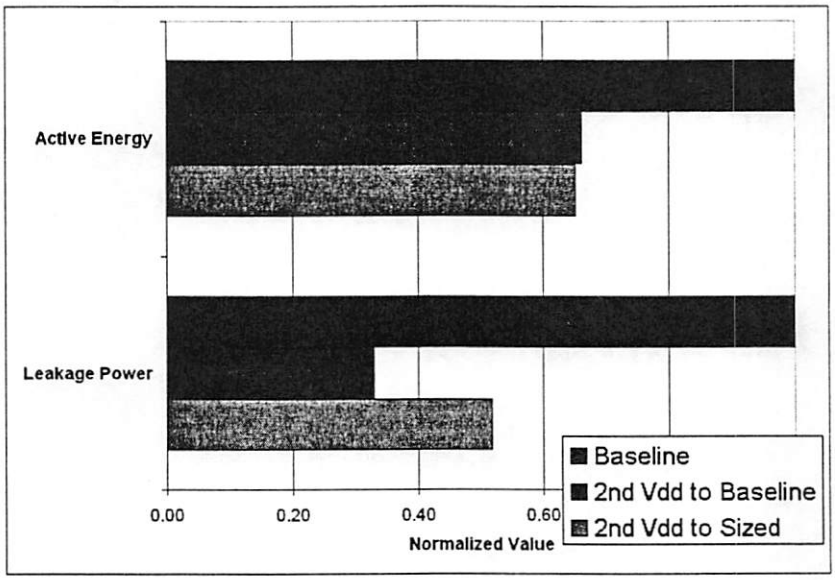


Figure 5.6 Effect of dual-supply (V_{DDL} = 0.8V) on power/energy.

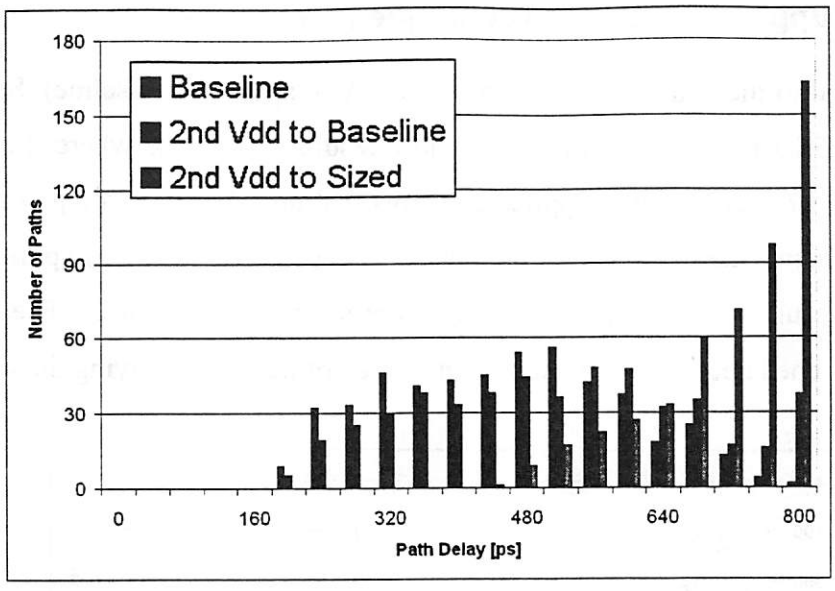


Figure 5.7 Effect of dual-supply (V_{DDL} = 1.0V) on path-delay distribution.

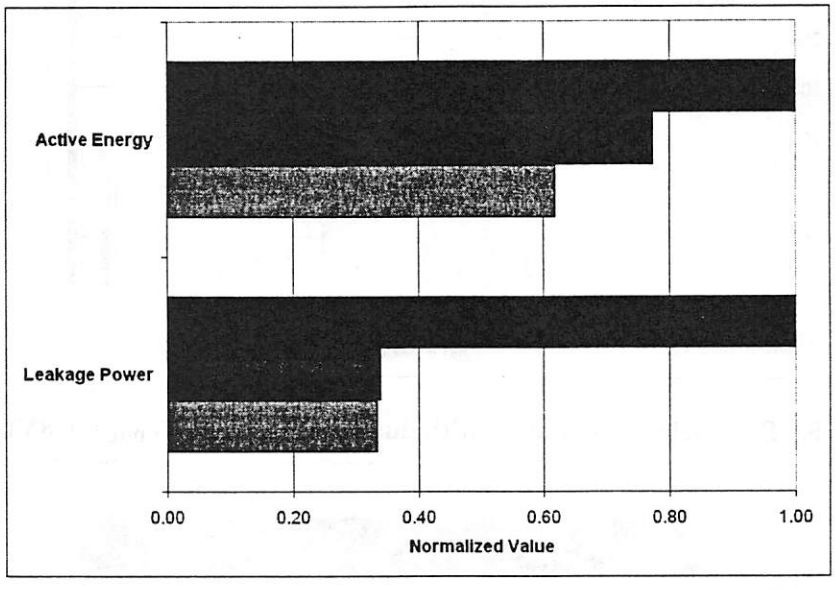


Figure 5.8 Effect of dual-supply (V_{DDL} = 1.0V) on power/energy.

5.3.2 Low Supply Voltage of 1.0V

See Figure 5.7 for the path-delay distribution and Figure 5.8 for the power/energy figures. These results underscore the importance of choosing the correct value for V_{DDL}.

5.4 Combination of Techniques

To determine if the techniques of dual-supply, dual-threshold and sizing are cumulative, they were applied in conjunction with each other.

5.4.1 Dual-Supply As Primary Technique (V_{DDL}= 0.8V)

Sizing was added to the dual-V_{DD} design of 0 (dual-V_{DD} applied to baseline). See Figure 5.9 and Figure 5.10 for results. Comparing these results with those where the baseline was sized and *then* dual-V_{DD} was applied confirms that dual-V_{DD} is more powerful than sizing for both active energy and leakage power. A second V_T was then applied to this design, which resulted in a small additional decrease in leakage power. The effect of dual-V_T was lessened because many more paths were critical after applying dual-V_{DD} and

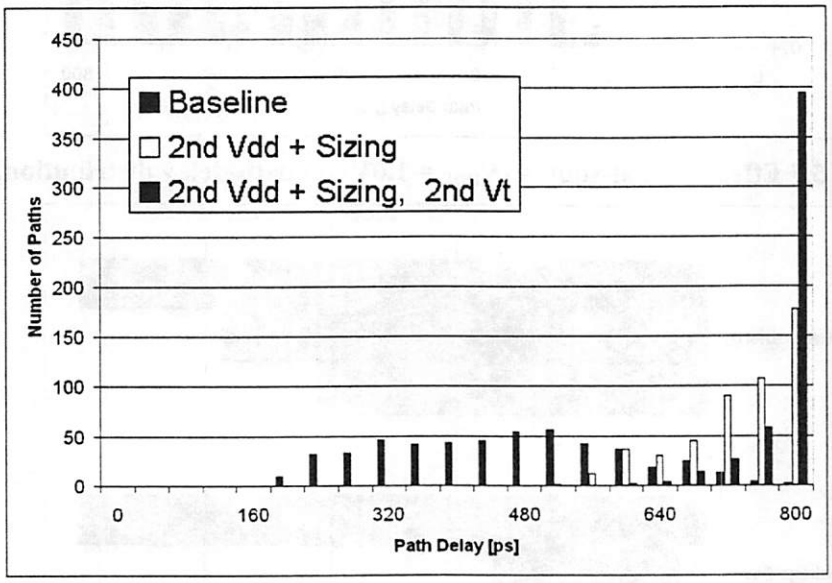


Figure 5.9 Path-delay distribution with dual-V_{DD} primary, (V_{DDL} = 0.8V).

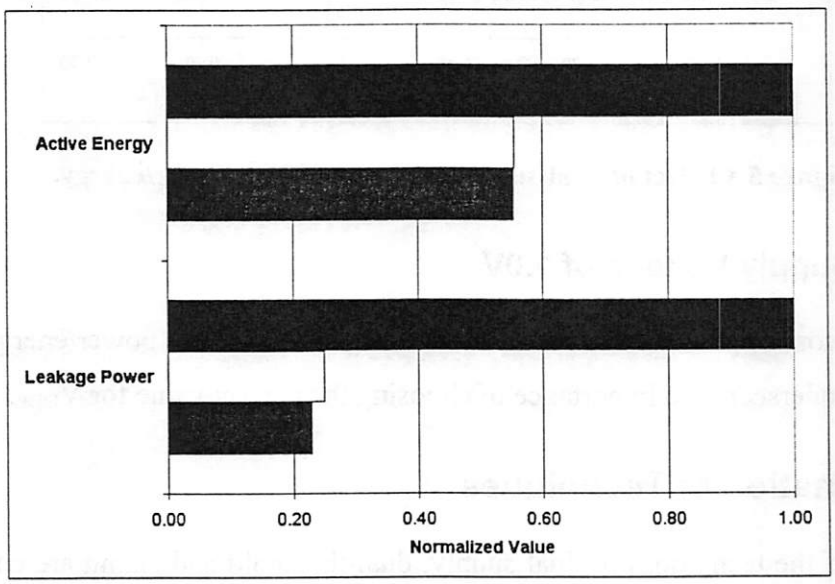


Figure 5.10 Power/energy with combination of techniques, dual-V_{DD} primary, (V_{DDL} = 0.8V).

sizing and could not absorb the large delay penalty of high-V_T cells.

5.4.2 Dual-Supply As Primary Technique (V_{DDL} = 1.0V)

The same procedures as in 5.2 (page26) were followed with a low supply voltage of 1.0V. With the higher second supply, leakage power performance was better than with a V_{DDL} of 0.8V, see Figure 5.11 and Figure 5.12, but active energy performance was worse. Also, with the higher V_{DDL}, the advantage of dual-supply over sizing was not as clear.

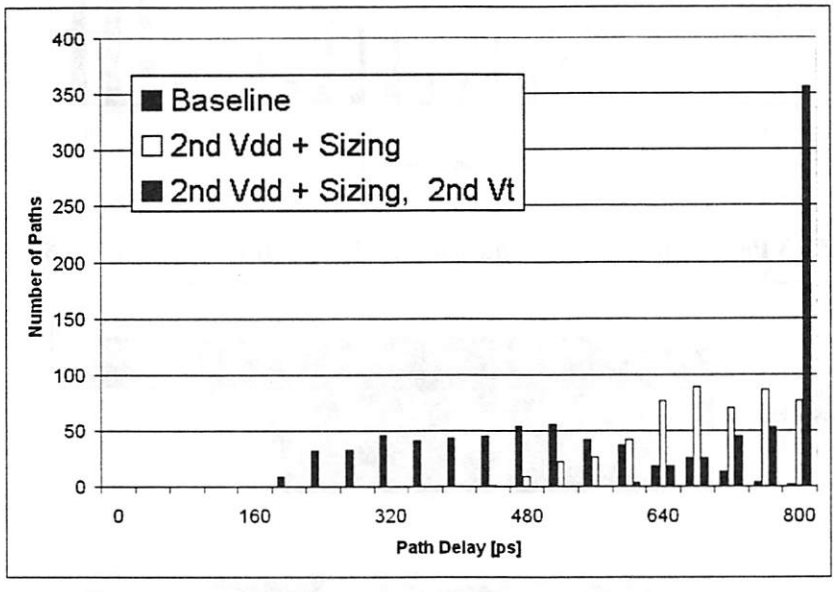


Figure 5.11 Path-delay distribution with dual-V_{DD} primary, (V_{DDL} = 1.0V).

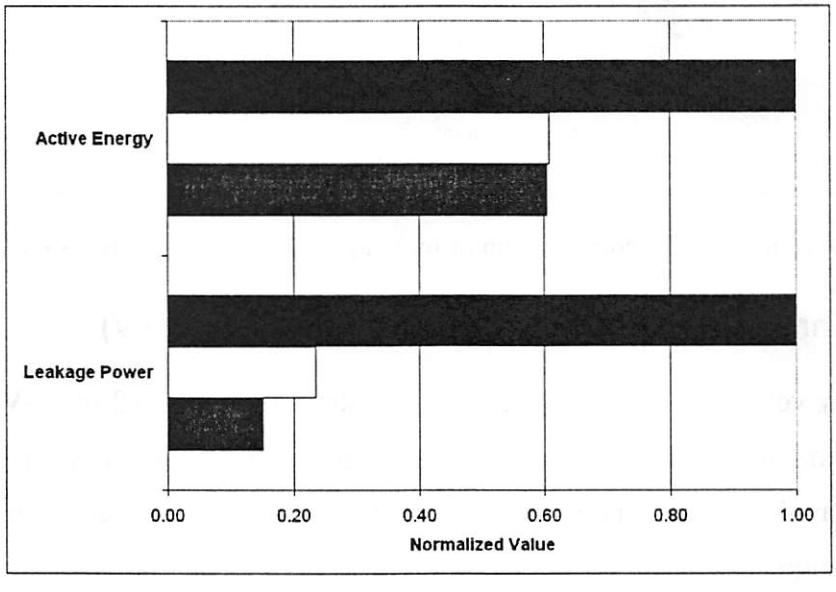


Figure 5.12 Power/energy with combination of techniques, dual-V_{DD} primary (V_{DDL} = 0.8V).

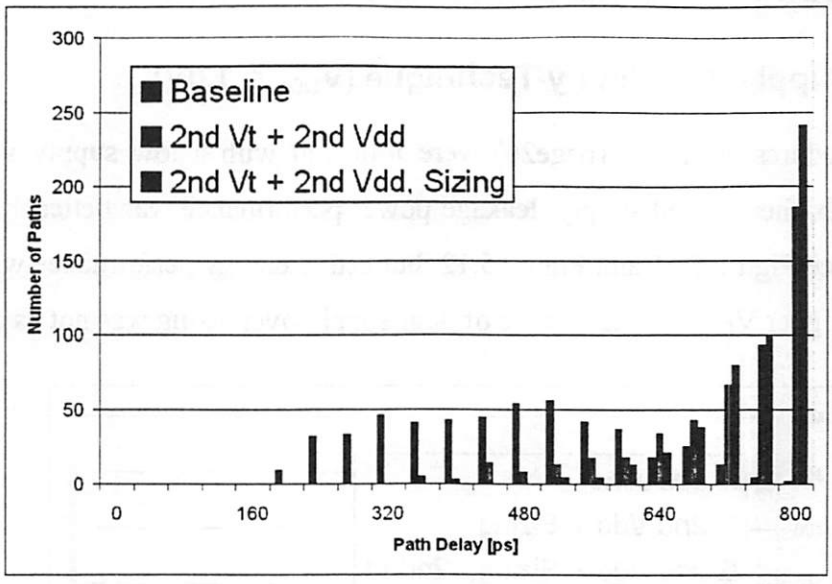


Figure 5.13 Path-delay distribution with dual-V_T primary, (V_{DDL} = 0.8V).

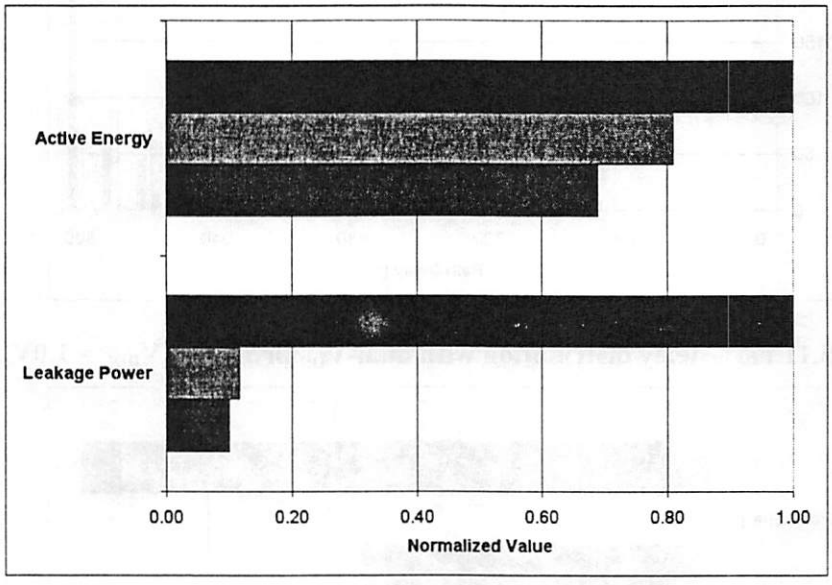


Figure 5.14 Power/energy with combination of techniques, dual-V_T primary, (V_{DDL} = 0.8V).

5.4.3 Dual-Threshold As Primary Technique (V_{DDL} = 0.8V)

A second supply voltage was added to the dual-threshold design of 5.2 (dual-V_T applied to baseline). Sizing was then used to further reduce power consumption. See Figure 5.13 for the path-delay distribution and Figure 5.14 for the power/energy figures. The benefits of all three techniques were compounded.

5.4.4 Dual-Threshold As Primary Technique (V_{DDL} = 1.0V)

As in 5.4.3 (page 32), dual-supply and then sizing were applied to the dual-threshold design. Here, the higher value of V_{DDL}, 1.0V, was used for dual-supply. The results did not vary from V_{DDL} of 0.8V.

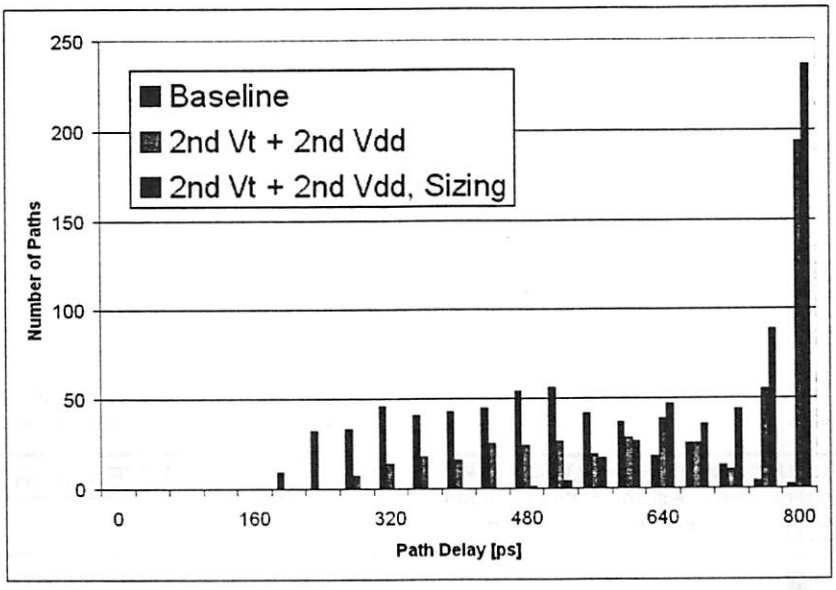


Figure 5.15 Path-delay distribution with combination of techniques, dual-V_T primary, (V_{DDL}=1.0V).

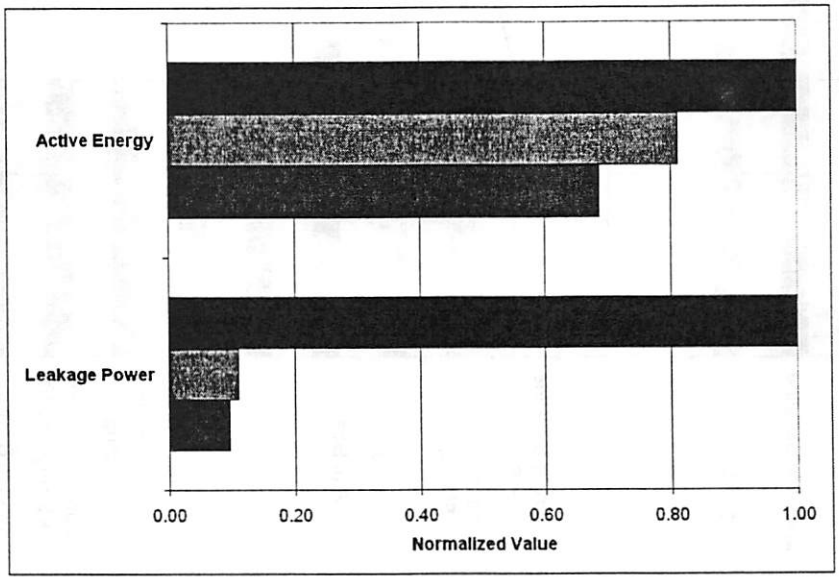


Figure 5.16 Power/energy with combination of techniques, dual-V_T primary, (V_{DDL}=1.0V).

5.5 Summary of Results

In total, fifteen different techniques or combination of techniques were applied to the baseline design with varying results. Figure 5.17 summarizes the normalized active energy values across all applied techniques, while Figure 5.18 (next page) summarizes the leakage power. In these figures, V_{DD2} refers to a low supply of 0.8V and V_{DD3} refers to a low supply of 1.0V.

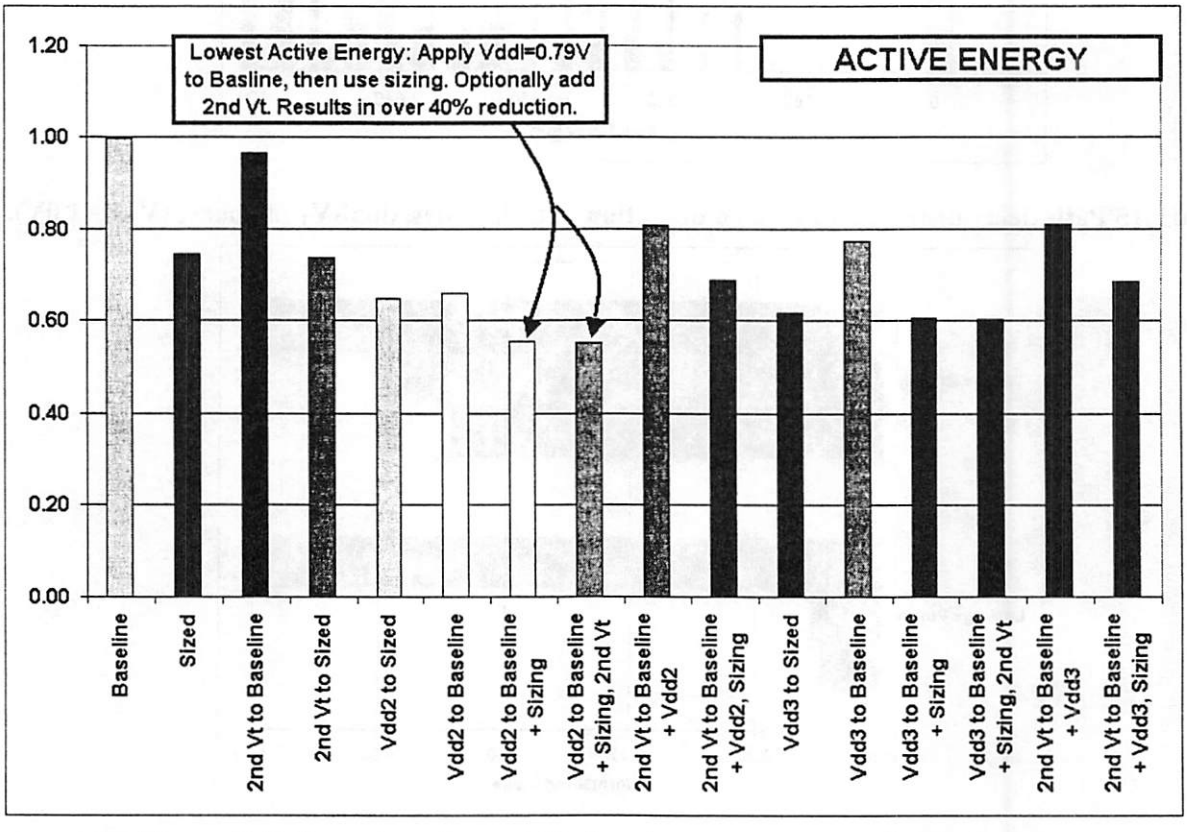


Figure 5.17 Summary of normalized active energy across all techniques.

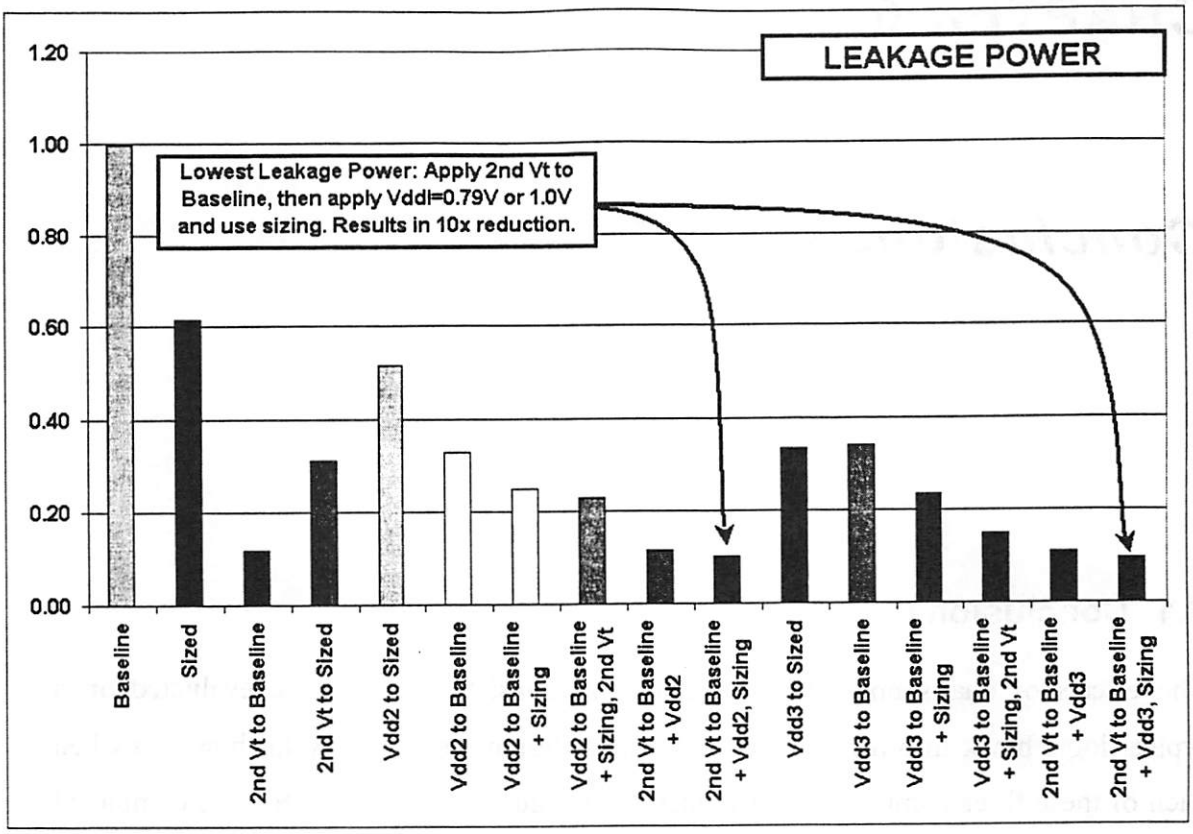


Figure 5.18 Summary of normalized leakage power across all techniques.

CHAPTER 6

Conclusions

6.1 Conclusions

The effects of dual-supply, dual-threshold and transistor sizing were evaluated on a typical logic block in order to gain a consistent design methodology for how and when each of these three common power reduction techniques should be used. The completed experimentation shows that energy savings from these three base techniques can be compounded through proper combination for additional benefit. Table 6.1 summarizes the best cases for active and leakage power reduction.

Table 6.1 Summary of Best Results.

Technique	Normalized Active Energy	Normalized Leakage Power
Lowest Active Energy		
Applied dual-V_{DD} (V_{DDL}=0.8V), sizing, dual-V_T	0.55	0.23
Applied dual-V_{DD} (V_{DDL}=0.8V), sizing	0.55	0.25
Applied dual-V_{DD} (V_{DDL}=1.0V), sizing, dual-V_T	0.60	0.15
Lowest Leakage Power		
Applied dual-V_T , dual-V_{DD} (V_{DDL}=0.8V), sizing	0.69	0.10
Applied dual-V_T , dual-V_{DD} (V_{DDL}=1.0V), sizing	0.69	0.10
Applied dual-V_T , dual-V_{DD} (V_{DDL}=1.0V)	0.81	0.11

There exists an optimal energy for a given block. However, the techniques of dual-supply, dual-threshold and sizing are typically applied sequentially. Depending on the order of application, different results are obtained for leakage and active power.

Overall power is dependent on many factors, including switching activity and the process. Depending on the activity of a logic block, a different emphasis should be placed on the techniques used. For high α, dual-supply should be the first technique applied, followed by transistor sizing and then dual-threshold, only if it does not impact the active power (this will depend on the value of V_{DDL}). However, if leakage power is the chief concern (low-α), dual-threshold takes precedence over the other techniques, followed by dual-supply and then transistor sizing.

Relative power savings are fairly general, but absolute numbers depend on the type of logic block. In this analysis, the starting point was a logic block with all paths sized for maximum speed with all low-V_T transistors. This presents an over design, but is common in today's designs. Presence of the large load at the output is also common. Lowering the supply voltage on the output load using the CVS technique is more effective than sizing at reducing power in the load. Without such substantial loading, dual-supply may not be superior to downsizing. In logic blocks with a large amount of reconvergent fanout, such as adders, the CVS method of dual-supply may not be effective. ECVS can be considered as a possible alternative.

Finally, the large power savings shown in this study should motivate EDA support for design environments that combine these techniques, particularly in the area of multiple-supply voltages.

6.2 Future Work

To enforce the credibility of these results and conclusions, these findings must be applied to practical circuits, rather than only a general logic block. The final step would be fabrication of test chips. This goal is currently being hindered due to lack of CAD support for dual-supply synthesis. Hopefully, the work presented in this thesis will stimulate the development of such a tool to facilitate the completion of this project.

The use of dual-supplies is an area that also needs further consideration, including second supply generation, layout and packaging. Variations in supplies, thresholds and sizing need to be considered, as well.

CHAPTER 7

References

[Burd00] T. Burd, T. Pering, A. Stratakos and R. Broderson, "A Dynamic Voltage Scaled Microprocessor System," *2000 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp 294-295, February 2000.

[Chandrakasan92] A. Chandrakasan, S. Sheng and R. Broderson, "Low-Power CMOS Digital Design," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 4, pp 473-484, April 1992.

[De00] V. De et al, "Techniques for Leakage Power Reduction," *Design of High-Performance Microprocessor Circuits*, IEEE Press, NJ, 2001, pp46-62.

[Edmonson00] J. Edmondson, "Impact of Physical Technology on Architecture," *Design of High-Performance Microprocessor Circuits*, IEEE Press, NJ, 2001, pp3-24.

[Hamada01] M. Hamada, Y. Ootaguro and T. Kuroda, "Utilizing Surplus Timing for Power Reduction," *Proceedings of the IEEE 2001 Custom Integrated Circuits Conference*, pp 89-92, May 2001.

[Holloway97] T. Holloway et al, "0.18μm CMOS Technology for High-Performance, Low-Power, and RF Applications," *Symposium on VLSI Technology Digest of Technical Papers*, pp 13-14, 1997.

[Ishihara02] F. Ishihara and B. Nikolic, "Level-Converting Flip-Flops for Dual-Supply Systems," To be published, 2002.

[Kao00] T. Kao and A. Chandrakasan, "Dual-Threshold Voltage Techniques for Low-Power Digital Circuits," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 7, pp 1009-1018, July 2000.

[Kato00] N. Kato et al, "Random Modulation: Multi-Threshold-Voltage Design Methodology in Sub-2V Power Supply CMOS," *IEICE Transactions on Electronics*, vol. E83-C, no.11, pp 1747-1754, November 2000.

[Kuroda98] T. Kuroda, "Low-Power CMOS Circuit Design by Means of Supply-Voltage and Threshold-Voltage Control," Ph.D. Dissertation, University of Tokyo, December 1998.

[Meindl95] J. D. Meindl, "Low power Microelectronics: Retrospect and Prospect," *Proceedings of the IEEE*, Vol. 83, No. 4, pp.619, 1995.

[Miyazaki02] M. Miyazaki, J. Kao and A. Chandrakasan, "A 175mV Multiply-Accumulate Unit using an Adaptive Supply Voltage and Body Bias (ASB) Architecture," *2002 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp 58-59, February 2002.

[Nikolic01] B. Nikolic, EE 241 Lecture Slides, University of California, Berkeley, Spring 2001.

[Rabaey96] J. Rabaey, *Digital Integrated Circuits: A Design Perspective*, Prentice Hall, 1996.

[Sakurai90] T. Sakurai and R. Newton, "Alpha-Power Law MOSFET Model and Its Applications to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp 584-594, April 1990.

[Thompson97] S. Thompson, I. Young, J. Greason and M. Bohr, "Dual Threshold Voltages and Substrate Bias: Keys to High Performance, Low Power, 0.1μm Logic Designs," *1997 Symposium on VLSI Technology Digest of Technical Papers*, pp 69-70, 1997.

[Usami00] K. Usami and M. Igarashi, "Low-Power Design Methodology and Applications Utilizing Dual Supply Voltages," *Proceedings of the Asia and South Pacific Design Automation Conference 2000*, pp 123-128, January 2000.

[Usami95] K. Usami and M. Horowitz, "Clustered Voltage Scaling for Low-Power Design," *International Symposium on Low Power Design*, pp 3-8, April 1995.

[Usami97] K. Usami et al, "Automated Low-Power Technique Exploiting Multiple Supply Voltages Applied to a Media Processor," *Proceedings of the IEEE 1997 Custom Integrated Circuits Conference*, pp 131-134, May 1997.

[Yamashita00] T. Yamashita et al, "A 450MHz 64b RISC Processor Using Multiple Threshold Voltage CMOS," *2000 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp 414-415, February 2000.

APPENDIX A

Preliminary Study Gate Models

VT/VDD	Cells	Delay [ps]	Active Energy [J]	Leakage Power [W]
For Use with Single/Dual/Triple-V_{DD}				
HS 1.8V	INV	19.94	5.26E-15	9.23E-10
	NAND2	24.50	5.05E-15	2.10E-11
	NAND3	50.70	1.79E-15	6.70E-12
	NAND4	47.20	1.92E-15	4.27E-10
	NOR2	33.48	1.23E-14	1.00E-09
	NOR3	41.33	1.17E-14	1.27E-11
	NOR4	86.56	4.06E-15	4.94E-12
LL 1.8V	INV	24.50	5.05E-15	2.10E-11
	NAND2	50.70	1.79E-15	6.70E-12
	NAND3	47.20	1.92E-15	4.27E-10
	NAND4	33.48	1.23E-14	1.00E-09
	NOR2	41.33	1.17E-14	1.27E-11
	NOR3	86.56	4.06E-15	4.94E-12
	NOR4	59.26	4.39E-15	2.85E-10
For Use with Dual-V_{DD}				
HS 1.1V	INV	47.20	1.92E-15	4.27E-10
	NAND2	33.48	1.23E-14	1.00E-09
	NAND3	41.33	1.17E-14	1.27E-11
	NAND4	86.56	4.06E-15	4.94E-12
	NOR2	59.26	4.39E-15	2.85E-10
	NOR3	47.05	2.13E-14	1.43E-09
	NOR4	59.83	1.99E-14	1.69E-11
LL 1.1V	INV	50.70	1.79E-15	6.70E-12

VT/VDD	Cells	Delay [ps]	Active Energy [J]	Leakage Power [W]
	NAND2	47.20	1.92E-15	4.27E-10
	NAND3	33.48	1.23E-14	1.00E-09
	NAND4	41.33	1.17E-14	1.27E-11
	NOR2	86.56	4.06E-15	4.94E-12
	NOR3	59.26	4.39E-15	2.85E-10
	NOR4	47.05	2.13E-14	1.43E-09
For Use with Triple-V_{DD}				
HS 0.86V	INV	29.15	2.46E-15	5.88E-10
	NAND2	52.15	1.15E-15	2.37E-10
	NAND3	33.48	1.23E-14	1.00E-09
	NAND4	49.93	5.66E-15	3.74E-10
	NOR2	91.61	2.59E-15	1.73E-10
	NOR3	47.05	2.13E-14	1.43E-09
	NOR4	71.57	9.62E-15	5.17E-10
HS 1.24V	INV	52.15	1.15E-15	2.37E-10
	NAND2	33.48	1.23E-14	1.00E-09
	NAND3	49.93	5.66E-15	3.74E-10
	NAND4	91.61	2.59E-15	1.73E-10
	NOR2	47.05	2.13E-14	1.43E-09
	NOR3	71.57	9.62E-15	5.17E-10
	NOR4	133.72	4.34E-15	2.33E-10

APPENDIX B

Gate Models, Comprehensive Study

Cell	Delay [ps]		Active Energy [J[		Leakage Power [W]	Input Capacitance [fF]
	Y-intercept	Slope	Y-intercept	Slope		
INV	15.1	2.8795	4.50E-03	1.60E-03	6.44E-09	2.2
INV_LL	23.534	3.5338	3.30E-03	1.70E-03	5.49E-11	2
INV_V2	23.184	3.4961	1.90E-03	8.00E-04	7.97E-10	2.2
INV_LLV2	38.509	4.4164	1.90E-03	8.00E-04	7.19E-12	2
INV_V3	14.607	3.5564	2.00E-03	1.40E-03	1.87E-09	2.2
INV_LLV3	26.183	4.2701	1.60E-03	1.50E-03	1.60E-11	2
NAND2	23.231	2.4549	9.30E-03	1.60E-03	8.23E-09	2.5
NAND2_LL	35.414	3.1117	7.80E-03	1.70E-03	7.47E-11	2.3
NAND2_V2	34.85	2.6443	4.10E-03	7.00E-04	9.33E-10	2.5
NAND2_LLV2	53.635	3.6638	4.00E-03	7.00E-04	9.81E-12	2.3
NAND2_V3	27.717	2.3929	6.50E-03	1.10E-03	9.81E-12	2.5
NAND2_LLV3	43.07	3.04	5.90E-03	1.10E-03	1.96E-11	2.3
NAND3	32.227	2.2141	1.47E-02	1.60E-03	1.05E-08	2.9
NAND3_LL	47.413	2.9531	1.28E-02	1.60E-03	9.32E-11	2.5
NAND3_V2	47.507	2.5937	7.30E-03	8.00E-04	1.14E-09	2.9
NAND3_LLV2	71.595	3.6952	6.90E-03	8.00E-04	1.29E-11	2.5
NAND3_V3	38.145	2.3207	1.07E-02	1.30E-03	2.54E-09	2.9
NAND3_LLV3	56.588	3.088	9.60E-03	1.30E-03	2.49E-11	2.5
NAND4	42.075	2.1525	2.14E-02	1.60E-03	1.50E-08	3.2
NAND4_LL	60.49	2.9443	1.85E-02	1.60E-03	1.38E-10	2.8
NAND4_V2	61.95	2.4527	1.13E-02	8.00E-04	1.40E-09	3.2
NAND4_LLV2	92.675	3.5927	1.04E-02	8.00E-04	1.64E-11	2.8
NAND4_V3	49.779	2.2216	1.62E-02	1.20E-03	3.06E-09	3.2

Cell	Delay [ps]		Active Energy [J[		Leakage Power [W]	Input Capacitance [fF]
	Y-intercept	Slope	Y-intercept	Slope		
NAND4_LLV3	71.822	3.0719	1.43E-02	1.20E-03	3.09E-11	2.8
NOR2	27.474	2.168	1.43E-02	1.60E-03	1.17E-08	3.4
NOR2_LL	40.835	2.7186	1.35E-02	1.60E-03	1.00E-10	3.2
NOR2_V2	42.915	2.8117	6.50E-03	8.00E-04	1.75E-09	3.4
NOR2_LLV2	64.348	4.1008	6.20E-03	8.00E-04	7.74E-11	3.2
NOR2_V3	33.981	2.1243	1.01E-02	1.20E-03	4.21E-09	3.4
NOR2_LLV3	50.547	2.7296	9.70E-03	1.20E-03	2.53E-10	3.2
NOR3	45.021	2.4523	2.53E-02	1.60E-03	1.59E-08	4
NOR3_LL	62.455	3.2463	2.45E-02	1.60E-03	1.32E-10	3.9
NOR3_V2	67.91	3.1421	1.21E-02	7.00E-04	2.46E-09	4
NOR3_LLV2	97.609	4.7844	1.16E-02	7.00E-04	1.02E-10	3.9
NOR3_V3	54.867	2.2776	1.87E-02	1.10E-03	5.88E-09	4
NOR3_LLV3	75.126	3.1077	1.77E-02	1.10E-03	3.37E-10	3.9
XOR	81.677	1.9578	1.70E-02	9.00E-04	2.23E-08	4
XOR_LL	113.83	2.6027	1.61E-02	9.00E-04	1.81E-10	3.8
XOR_V2	125.96	3.0485	6.90E-03	4.00E-04	5.78E-09	4
XOR_LLV2	200.51	5.1558	6.60E-03	4.00E-04	4.95E-11	3.8
XOR_V3	96.402	2.4924	1.11E-02	7.00E-04	1.16E-08	4
XOR_LLV3	139.37	3.5705	1.08E-02	7.00E-04	9.64E-11	3.8
LD	0		3.39E-02		0.00E+00	175
LDV2	36		9.49E-02		0.00E+00	175
LDV3	0		4.38E-02		0.00E+00	175

APPENDIX C

Summary of Results

Approach	Active Energy [pJ]	Normalized Active Energy	Leakage Power [μW]	Normalized Leakage Power	Maximum Delay [ps]
Initial un-sized	214.86	0.73	45.86	0.60	1142.58
BASELINE	293.34	1.00	76.80	1.00	781.23
Sizing					
Sizing to baseline	219.05	0.75	47.38	0.62	781.23
Dual-Threshold					
Dual-V_T to baseline	283.63	0.97	9.08	0.12	781.23
Dual-V_T to sized	216.12	0.74	23.94	0.31	781.23
Dual-Supply, V_{DDL}=0.8V					
Dual-V_{DD} to baseline	194.17	0.66	25.32	0.33	781.23
Dual-V_{DD} to sized	190.92	0.65	39.68	0.52	781.23
Dual-Supply, V_{DDL}=1.0V					
Dual-V_{DD} to baseline	227.16	0.77	26.07	0.34	781.23
Dual-V_{DD} to sized	181.45	0.62	25.71	0.33	781.23
Combination of Techniques, Dual-Supply primary, V_{DDL}=0.8V					
Dual-V_{DD}, sizing	162.79	0.55	19.19	0.25	781.23
Dual-V_{DD}, sizing, dual-V_T	162.45	0.55	17.63	0.23	781.23
Combination of Techniques, Dual-Supply primary, V_{DDL}=1.0V					
Dual-V_{DD}, sizing	178.22	0.61	18.01	0.23	781.23
Dual-V_{DD}, sizing, dual-V_T	177.27	0.60	11.61	0.15	781.23

Approach	Active Energy [pJ]	Normalized Active Energy	Leakage Power [μW]	Normalized Leakage Power	Maximum Delay [ps]
Combination of Techniques, Dual-Threshold primary, V_{DDL}=0.8V					
Dual-V_T, dual-V_{DD}	237.58	0.81	8.95	0.12	781.23
Dual-V_T, dual-V_{DD}, sizing	201.87	0.69	7.80	0.10	781.23
Combination of Techniques, Dual-Threshold primary, V_{DDL}=1.0V					
Dual-V_T, dual-V_{DD}	237.91	0.81	8.54	0.11	781.23
Dual-V_T, dual-V_{DD}, sizing	201.35	0.69	7.44	0.10	781.23